

# Extropy: a complementary dual of entropy

Frank Lad \*

Giuseppe Sanfilippo <sup>†</sup>

Gianna Agrò <sup>‡</sup>

September 30, 2011

## Abstract

This article resolves a longstanding question in the axiomatisation of entropy as proposed by Shannon and highlighted in renewed concerns expressed by Jaynes. We introduce a companion measure of a probability distribution that we suggest be called the *extropy* of the distribution. The entropy and the extropy of an event distribution are identical. However, this identical measure bifurcates into distinct measures for any quantity that is not merely an event indicator. As for entropy, the maximum extropy distribution is also the uniform distribution. We display several theoretical and geometrical properties of the proposed extropy measure, discussing in detail the difference between its assessment of a refined probability distribution and the axiom that characterises the Shannon entropy in this regard. This is what resolves the concerns of Shannon and Jaynes. In a discrete context, the extropy measure is approximated by a variant of Gini’s index of heterogeneity when the maximum probability mass is small. This is related to the “repeat rate” of a mass function as studied by Turing and Good. The continuous analogue of extropy turns out to equal the negative integral of the square of the density function. We conclude with a consideration of a rescaled measure of extropy which identifies it as the dual of entropy. The structure of the duality suggests a general theory of complementary distributions.

**Key Words:** entropy; extropy; Gini index of heterogeneity; repeat rate; duality; proper scoring rules

## 1 Motivation, scope, and background

The *entropy* measure of a probability distribution has had myriad useful applications in information sciences since its full-blown introduction in the extensive article of Shannon (1948).

---

\*Department of Mathematics and Statistics, *University of Canterbury*, Christchurch, New Zealand, F.Lad@math.canterbury.ac.nz

<sup>†</sup>Dipartimento di Scienze Statistiche e Matematiche, *University of Palermo*, Italy, giuseppe.sanfilippo@unipa.it

<sup>‡</sup>Dipartimento di Scienze Statistiche e Matematiche, *University of Palermo*, Italy, gianna.agro@unipa.it

Prefigured by its usage in thermodynamics by Boltzmann and Gibbs, entropy has subsequently bloomed as a showpiece in theories of communication, coding, probability and statistics. So widespread is its application and advocacy, it is with due respect that we propose this measure has a natural complement which merits recognition and comparison, perhaps in many realms of its current application ... the measure of *extropy*.

In this article we display several intriguing properties of this information measure which identify extropy as both a complement and a dual to entropy. Not only does its recognition resolve a fundamental question that has surrounded Shannon's measure since its very inception, but it also provides links to other notable information measures whose relation to entropy have not been recognised. Our presentation of the properties of extropy is meant to air the results for readers relatively familiar with the Shannon entropy theory. We shall follow his notation and extend it. Discussion of our specific applied motivation for generating these results, and their relation to the theory of proper scoring rules would be distracting as an introduction.

Suppose  $X$  is a quantity with a finite discrete realm of possibilities  $\{x_1, x_2, \dots, x_N\}$ . If a probability distribution over the partition of events  $(X = x_1), (X = x_2), \dots, (X = x_N)$  is composed of the vector probability masses  $\mathbf{p}_N = (p_1, p_2, \dots, p_N)$ , the Shannon entropy measure denoted by  $H(X)$  or  $H(\mathbf{p}_N)$  equals  $-\sum_{i=1}^N p_i \log(p_i)$ . The complementary measure we propose as extropy, and denote here by  $J(X)$  or  $J(\mathbf{p}_N)$ , equals  $-\sum_{i=1}^N (1 - p_i) \log(1 - p_i)$ . As is entropy, extropy is interpreted as a measure of the amount of uncertainty represented by the distribution for  $X$ . The results that conclude this note will suggest a different location and scaling for extropy inhering in the alternative measure  $J^*(\mathbf{p}_N) = (N - 1)^{-1} J(\mathbf{p}_N) + \log(N - 1)$ . With this scaling, the extropy measure can be recognised formally as the dual of entropy. Moreover, its function value for a mass function equals the entropy of a complementary distribution which shall be identified in our Section 6.

Shannon's (1948) original and fairly exhaustive investigation of entropy characterised it as the unique (up to location and scale transformations) measure  $H(\cdot)$  of a mass function  $\mathbf{p}_N$  over a partition of events that satisfies three properties:

- i.)  $H(p_1, p_2, \dots, p_N)$  is continuous in each of its arguments;
- ii.)  $H(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})$  is a monotonic increasing function of the partition size,  $N$ ; and
- iii.)  $H(tp, (1 - t)p, (1 - p)) = H(p, 1 - p) + p H(t, 1 - t)$ .

The article of Renyi (1961) presented alternative characterisations of entropy due to Fadeev and himself. These involved alternating these axioms with various properties of the Shannon measure, such as its invariance with respect to permutations of its arguments, and its achieved

maximum occurring at the uniform distribution.

Shannon's third axiom implies that the entropy in a joint distribution for two quantities equals the entropy in the marginal distribution for one of them plus the expectation for the entropy in the conditional distribution for the second given the first:

$$H(X, Y) = H(X) + \sum_{x_i} P(X = x_i) H(Y|X_i = x_i) . \quad (1)$$

Indeed the appeal of this result was a motivation favouring Shannon's choice of his axiom iii.

In his original extensive article, Shannon (p. 50 in the 1949 reprint) slighted his own characterisation theorem for entropy, noting that it was in no way necessary for the larger theory of communication he was developing. He viewed it merely as lending plausibility to some subsequent definitions. He considered the real justification of the three axioms of entropy as residing in their desirable useful implications. In particular, the implication that the joint entropy in two quantities equals the entropy in one plus the expected conditional entropy in the other given that one (our equation 1) was regarded as welcome substantiation for entropy as a reasonable measure of information.

The thoughtful discussion of Jaynes (2003, Section 11.3), who was a major contributor to the understanding of entropy and its importance, explicitly recognised the discussable open status of entropy's third characterising axiom. Having developed his discussion of probability some 350 pages without requiring it, he highlighted it as "really an additional assumption which we should have included in our list." He followed this statement with an "Exercise 11.1" which concludes with the injunction to "Carry out some new research in this field by investigating this matter; try either to find a possible form of the new functional equations, or to explain why this cannot be done."

As we read him, Jaynes clearly expected that a satisfactory motivation for the special status of entropy as a measure of information would be found, thinking that his "exercise" would be resolved with a solution explaining "why this cannot be done". In a direct sense, our construction and analysis of the extropy measure shows the exercise to be solved rather by the exhibition of a "new functional equation", providing an alternative to Shannon's third axiom. Jaynes' expectations regarding this matter led him, we believe, to one of his rare overstatements of the status of entropy as a *unique* measure of information. He wrote (2003, p. 350) "We have shown that use of the measure (Shannon entropy) is a *necessary* condition for consistency", and further conjectured "that any other choice of 'information measure' will lead to inconsistencies if carried far enough". We only remark that to be precise, what was shown is that Shannon's definition of entropy is necessary for consistency *with the third proposed axiom*. Concerns with a foundational

establishment of the uniqueness of entropy were also aired by Kolmogorov (1956, p. 105).

Despite his expectations, Jaynes was not convinced that an adequate foundation for the *uniqueness* claims of entropy as an information measure had been found. He concluded that long section of his book by writing (p. 351) “Although the above demonstration appears satisfactory mathematically, it is not yet in completely satisfactory form conceptually. The functional equation (Shannon’s third axiom) does not seem quite so intuitively compelling as our previous ones did. In this case, the trouble is probably that we have not yet learned how to verbalize the argument leading to (axiom iii) in a fully convincing manner. Perhaps this will inspire others to try their hand at improving the verbiage that we used just before writing (axiom iii).” We believe that a person of Jaynes’ imagination and insight would have been pleased with our surprising resolution of his dissatisfaction. In tandem with Shannon’s entropy measure denoted by  $H(\cdot)$ , we appropriately denote our extropy measure by  $J(\cdot)$ .

## 2 The characterisation of extropy

**Context:** Suppose that the possible values of an unknown but observable quantity  $X$  are the numbers in the *realm*  $\Re(X) = \{x_1, x_2, \dots, x_N\}$ . The vector  $\mathbf{p}_N = (p_1, p_2, \dots, p_N)$  denotes an associated probability mass function asserted for  $X$  over the event partition  $\{(X = x_1), (X = x_2), \dots, (X = x_N)\}$ . We recall

**Definition 1:** The *entropy* in  $X$  or in  $\mathbf{p}_N$  equals

$$H(X) = H(\mathbf{p}_N) = - \sum_{i=1}^N p_i \log(p_i) , \quad (2)$$

and we introduce

**Definition 2:** The *extropy* in  $X$  or in  $\mathbf{p}_N$  equals

$$J(X) = J(\mathbf{p}_N) = - \sum_{i=1}^N (1 - p_i) \log(1 - p_i). \quad (3)$$

**Result 1:** If  $N = 2$ , so  $X$  is merely an event, then  $H(X) = J(X)$ ,

$$\text{for } H(\mathbf{p}_2) = p_1 \log(p_1) + (1 - p_1) \log(1 - p_1) = J(\mathbf{p}_2).$$

However,  $H(\mathbf{p}_N) > J(\mathbf{p}_N)$  as long as  $\mathbf{p}_N$  contains three or more non-zero components.

When  $N = 3$  it is no longer necessary that  $H(\mathbf{p}_3)$  equals  $J(\mathbf{p}_3)$ . In fact, these will be equal only for mass functions  $\mathbf{p}_3$  that have one component equal to 0. (In this case the distribution has only two non-zero components which sum to 1, just as a distribution when  $N = 2$ . By convention,  $p_i \log(p_i) \equiv 0$  when  $p_i = 0$ , preserving continuity.) Figure 1 displays the range

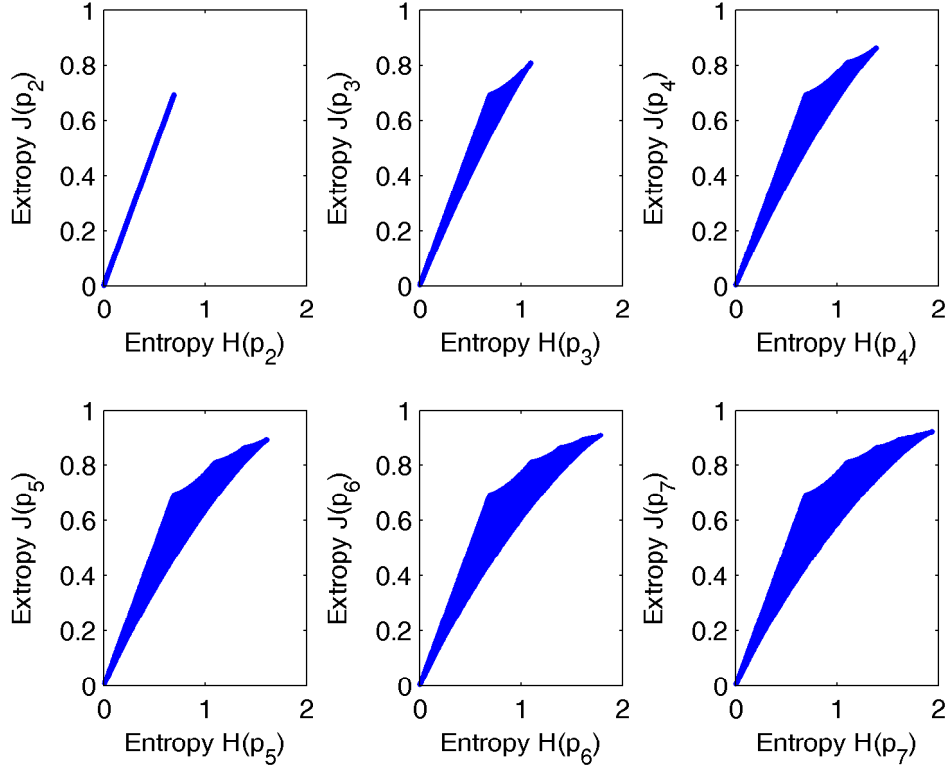


Figure 1: The range of (entropy, extropy) pairs  $(H(\cdot), J(\cdot))$  corresponding to all distributions within the unit-simplices of dimension 1 through 6. The realms of the quantities they assess have sizes  $N = 2$  through 7.

of possibilities for the (entropy, extropy) pairs for probability mass functions within the unit-simplices of Dimension 1 through 6. This range expands regularly as the size of  $\mathfrak{R}(X)$  increases from 2 to 7; for a unit-simplex of Dimension  $K$  contains the unit-simplex of dimension  $(K - 1)$  as one of its “faces”. (The added dimension of possibility may be assessed with probability zero, so in this case the associated distributions would have the same range for  $H$  and  $J$  as the lower dimension.) An algebraic proof that  $H(\mathbf{p}_N) \geq J(\mathbf{p}_N)$  is submitted as an appendix.

Notice particularly that the range of possible (entropy, extropy) pairs is *not* convex. As viewed across the six examples shown in Figure 1, the range exhibits convex scallops along its upper boundary: there are  $(N-2)$  scallops and one flat edge along its upper boundary for the unit-simplex of dimension  $(N-1)$ . The flat edge as the northwest boundary is the line defined by  $H(p, 1-p) = J(p, 1-p)$ , running in the southwest to northeast direction from  $(0,0)$  to  $(-\log(.5), -\log(.5))$ . The lower boundary of the range of pairs is a single concave scallop, ruling its own interior out of the range.

**Result 2.**  $J(X)$  satisfies axioms Shannon’s axioms i and ii.

The function  $J(\cdot)$  is evidently continuous in its arguments, and

$$J(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}) = -N (1 - \frac{1}{N}) \log(1 - \frac{1}{N}) = -(N-1) \{ \log(N-1) - \log(N) \}$$

is a monotonic increasing function of  $N$ .

As to other touted properties of entropy, extropy shares many of them. For example the extropy measure is obviously permutation invariant. Moreover, for any size of  $N$ , the maximum extropy distribution is the uniform distribution. We prove this as follows. Let  $L(\mathbf{p}_N, \lambda)$  be the Lagrangian expression for the extropy of  $\mathbf{p}_N$  subject to the constraint  $\sum p_i = 1$ . Then

$$L(\mathbf{p}_N, \lambda) = - \sum_{i=1}^N (1 - p_i) \log(1 - p_i) + \lambda (1 - \sum_{i=1}^N p_i) \quad (4)$$

with  $N$  partial derivatives of the form  $\frac{\partial L}{\partial p_i} = \log(1 - p_i) + 1 - \lambda$ . Setting each of these equal to 0 yields  $N$  equations of the form  $\lambda = 1 + \log(1 - p_i)$ . These  $N$  equations, together with  $\frac{\partial L}{\partial \lambda} = 0$ , ensure that all the  $p_i$  are equal, and thus they must each equal  $1/N$ . The bordered Hessian determinants for the matrix of cross-partial derivatives alternate in sign, assuring that  $L(\cdot, \cdot)$  achieves a maximum at this solution.

The scale of the maximum entropy measure is unbounded as  $N$  increases, since  $H(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}) = \log(N)$ . In contrast, the scale of the maximum extropy is bounded by 1, for  $J(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}) = -(N-1) \log\{(N-1)/N\}$ . The limit of 1 can be determined by observing that

$$\lim_{N \rightarrow \infty} -(N-1) \log\left(\frac{N-1}{N}\right) = \lim_{N \rightarrow \infty} -\log\left(1 - \frac{1}{N}\right)^{N-1} = -\log(e^{-1}) = 1.$$

It is interesting now to examine precisely why extropy does *not* satisfy Shannon's third axiom for entropy, and how it does behave with respect to measuring the refinement of a probability distribution.

**Result 3.**  $J(tp, (1-t)p, 1-p) = J(p, 1-p) + \Delta(p, t)$ , where  
 $\Delta(p, t) = (1-p)\log(1-p) - (1-t)p\log(1-tp) - \{1 - (1-t)p\}\log\{1 - (1-t)p\}.$

This result follows easily from the definition of  $J(\mathbf{p}_3)$ . It is most easily interpreted visually. The right panel of Figure 2 displays the extropy  $J(p, 1-p)$  along with the difference between the extropies  $J(tp, (1-t)p, 1-p)$  and  $J(p, 1-p)$  according to Result 3, while the left panel displays the difference between the entropies  $H(tp, (1-t)p, 1-p)$  and  $H(p, 1-p)$  according to Shannon's axiom iii. Each difference is shown as a function of  $p \in [0, 1]$  for the four values of  $t = .1, .2, .3$ , and  $.5$ . Either difference function for any value of  $t$  would be the same as the function for the value  $(1-t)$ .

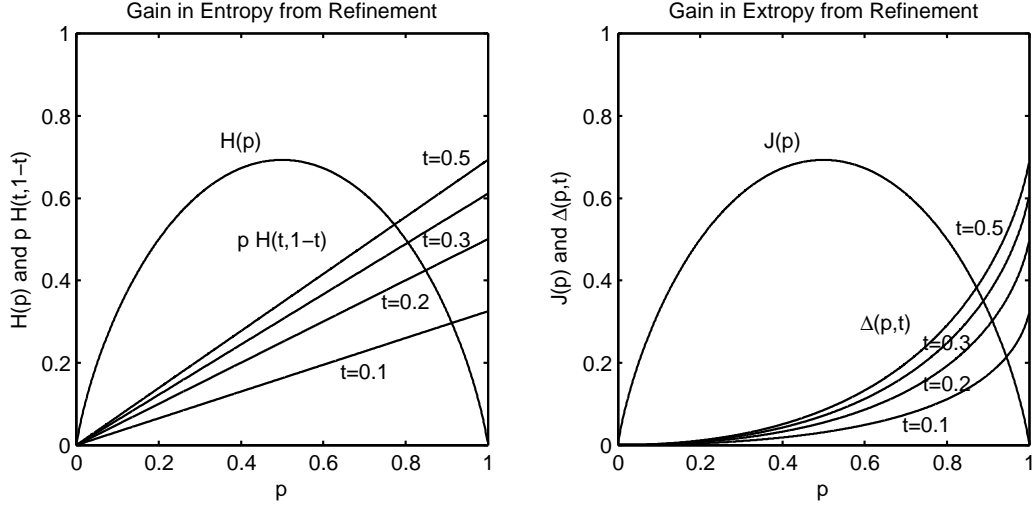


Figure 2: Entropy and extropy for a refined distribution  $(tp, (1-t)p, 1-p)$  both equal the extropy or entropy for the base probabilities  $(p, 1-p)$  plus an additional component.

According to Shannon's axiom iii, the entropy for the refined distribution  $(tp, (1-t)p, 1-p)$  increases linearly with  $p$  at the rate of the entropy in the refining split factor,  $H(t, 1-t)$ . In contrast, the extropy of the refined distribution increases at an increasing rate as a function of  $p$ . For small values of  $p$ , the extropy of the refined distributions increase more slowly with  $p$  than does entropy, while for large values of  $p$  it increases more quickly. As  $p$  increases to 1, the increases in the entropy and extropy of the refined distribution become equal, for any  $t \in [0, 1]$ . This is a result of the fact that when  $p = 1$ , the refined distribution is virtually a binary distribution  $(t, 1-t, 0)$ , for which entropy and extropy are equal. The distribution that is being refined would then be a degenerate distribution representing certainty.

As a gauge of the increase in uncertainty provided when a distribution is refined, this non-linear feature of the extropy measure is appealing. Refining a larger probability with a splitting factor of size  $t$  may well be considered to increase the amount of information at a greater rate than refining a smaller probability by this same factor. This is a natural feature of the extropy measure. Replacing Shannon's axiom iii with our Result 3 would complete the characterisation of extropy.

**Result 4.**  $H(\mathbf{p}_N) + J(\mathbf{p}_N) = \sum_{i=1}^N H(p_i, 1-p_i) = \sum_{i=1}^N J(p_i, 1-p_i).$

Thus,  $J(\mathbf{p}_N) = \sum_{i=1}^N H(p_i, 1-p_i) - H(\mathbf{p}_N).$

Symmetrically,  $H(\mathbf{p}_N) = \sum_{i=1}^N J(p_i, 1-p_i) - J(\mathbf{p}_N).$

The basic equation for the sum of  $H(\cdot)$  and  $J(\cdot)$  derives from simply summing the two com-

ponents of each of the  $H(p_i) = p_i \log(p_i) + (1 - p_i) \log(1 - p_i)$ . The resulting equation displays that the extropy of a distribution equals the difference between the sum of the entropies in the crudest partitions defined by the possible values of  $X$ , that is  $(X = x_i)$  and  $(X \neq x_i)$ , and the extropy in the finest partition they define:  $(X = x_1), (X = x_2), \dots$  and  $(X = x_N)$ . Moreover, since the entropy of any event equals the extropy of the event, this relation is symmetric in the functions  $H(\cdot)$  and  $J(\cdot)$ . It is apparent that the symmetric relation between entropy and extropy is fundamentally related to the refinement characteristics inherent in their third axioms. It also suggests a sense in which the extropy measure is a dual of entropy. This idea will be explored further in Section 5.

### 3 Isoentropy and Isoextropy contours in the unit-simplex

For the display that follows, we suppose that a quantity  $X$  has realm  $\mathfrak{R}(X) = \{1, 2, 3\}$ , and that these possibilities are assessed with a probability mass function  $\mathbf{p}_3$  in the unit-simplex  $\mathbf{S}^2$ . Figure 3Left displays some contours of constant entropy distributions in the 2-Dimensional unit-simplex ( $N = 3$ ), to compare with some contours of constant extropy distributions in Figure 3Right. These contours exhibit a geometrical sense in which the extropy and entropy measures of a distribution are complementary. Whereas entropy contours sharpen into the vertices of the simplex and flatten along the faces, the extropy contours sharpen into the midpoints of the faces and flatten into the vertices.

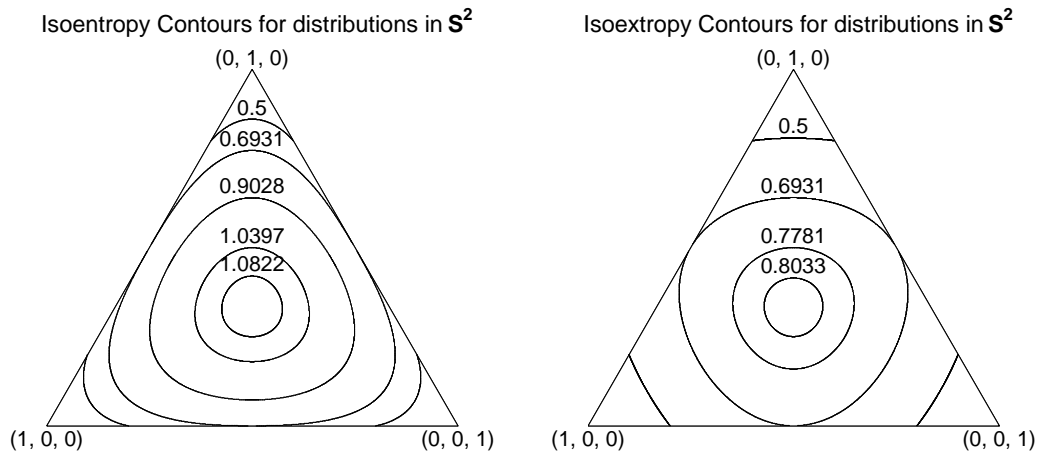


Figure 3: At left are contours of equal entropy distributions within the 2-D unit-simplex. At right are contours of equal extropy distributions.



## 4 The extropy measure of a continuous distribution

Devising the extropy measure of a continuous distribution admitting a density function yields a pleasant surprise. Shannon (1948, Section III.20) already proposed that the entropy measure  $-\sum p_i \log(p_i)$  has a continuous analogue in the measure  $-\int f(x) \log f(x) dx$  when the distribution function admits a continuous density. Kolmogorov (1956) concurred, with slight qualifying reservations. However, he also voiced concerns about the primacy of the entropy measure (p. 105), insisting that further analysis was needed. The analogical character of Shannon's entropy measure for a continuous density derives from its status as the limit of a linear transformation of the discrete entropy measure.

### 4.1 Shannon's continuous entropy, $-\int f(x) \log f(x) dx$

For the following simple exposition of Shannon's considerations, presume again that the realm of a quantity  $X$  is  $\{x_1, \dots, x_N\}$ , and that the values of  $x_1$  and  $x_N$  are fixed. For each larger value of  $N$ , presume that more elements are included uniformly in the interval between them. Define  $\Delta x \equiv (x_N - x_1)/(N - 1)$  for any specific  $N$ , and define  $f(x_i) \equiv p_i/\Delta x$ . In these terms, the entropy  $H(\mathbf{p}_N)$  can be expressed as

$$\begin{aligned} -\sum p_i \log(p_i) &= -\sum f(x_i) \Delta x \log\{f(x_i) \Delta x\} \\ &= -\sum f(x_i) \log\{f(x_i)\} \Delta x - \log\{\Delta x\} \end{aligned} \quad (5)$$

Thus,  $-\sum f(x_i) \log\{f(x_i)\} \Delta x$  is merely a location transform of the entropy  $-\sum p_i \log(p_i)$ , shifting only by the size of  $\log\{\Delta x\}$  which is finite for any  $N$ . The limit of the relocated entropy expression suggests the continuous analogue as  $-\int f(x) \log f(x) dx$ .

Shannon noted that this analagous measure loses the absolute meaning that the finite measure enjoys, because its value must be considered relative to an assumed standard of the coordinate system in which the value of the variable is expressed. (If the variable  $X$  were transformed into  $Y$ , then the continuous measure of the entropy is adjusted by the Jacobian of the transformation.) However, the continuous analogue retains its value as a comparative measure of the uncertainties contained in two densities because they would both be affected by the transformation in the same way.

### 4.2 Motivating the continuous extropy measure as $-\int f^2(x) dx$

At first sight, the extropy measure  $-\sum (1 - p_i) \log(1 - p_i)$  appears problematic: if each  $p_i$  were simply replaced by a density value  $f(x)$ , the measure would not be defined when  $f(x) > 1$ ,

which it may. However, the situation clarifies by expanding  $(1 - p_i) \log(1 - p_i)$  through three terms of its Maclaurin series with remainder:  $(1 - p_i) \log(1 - p_i) = -p_i + p_i^2/2 + r_i^3/6$  for some  $r_i \in (0, p_i)$ . Summing these expansion terms over  $i = 1, \dots, N$  shows that when the realm of possibilities for  $X$  increases (as a result of larger  $N$ ) in such a way that  $\Delta x \rightarrow 0$  and  $\max_{i=1}^N p_i$  decreases toward 0, the extropy measure becomes closely approximated by  $1 - \frac{1}{2} \sum_{i=1}^N p_i^2$ .

Following the same tack as for entropy in representing  $p_i$  by  $f(x_i)\Delta x$  suggests that for large  $N$  the analogue continuous extropy measure can be approximated by

$$\begin{aligned} 1 - \frac{1}{2} \sum_{i=1}^N p_i^2 &= 1 - \frac{1}{2} \sum f^2(x_i) (\Delta x)^2 \\ &= 1 - \frac{\Delta x}{2} \sum f^2(x_i) \Delta x . \end{aligned} \quad (6)$$

This approximation is merely a location and scale transformation of the term  $-\sum f^2(x_i)\Delta x$ . Thus, the limiting measure of extropy for a continuous density is well regarded as  $-\int f^2(x)dx$ .

The sum of the squares of probability masses (as well as the integral of the square of a density) has received attention for more than a century for a variety of reasons, but never in a direct relation to the entropy of a distribution. Rather, it has merely been considered an alternative measure of uncertainty. Good (1979) referred to this measure as the “repeat rate” of a distribution, developing an original idea of Turing. Gini (1912, 1939) had earlier proposed this measure as an “index of heterogeneity” of a discrete distribution, via  $1 - \sum_{i=1}^N p_i^2$ . We now find that in a discrete context, a rescaling of Gini’s index is an approximation to the extropy of a distribution when the maximum probability mass is small. In a continuous context, the negative expected value of a density function value is the analogue of the extropy measure of a distribution that we are proposing.

## 5 Rescaling extropy: a theory of complementary distributions

Let us return to the definition of extropy for a discrete probability mass function over a finite partition. We have noted in Section 2 how the scaling of our entropy measure  $H(\cdot)$  is different from that of our extropy measure  $J(\cdot)$ : the range of the former is unbounded as  $N$  increases, while that of the latter is bounded by 1.

Suppose we redefine extropy as extropy\* according to a location and scale transformation:

$$\begin{aligned} J^*(\mathbf{p}_N) &\equiv (N-1)^{-1} J(\mathbf{p}_N) + \log(N-1) \\ &= - \sum_{i=1}^N \frac{1-p_i}{N-1} \log\left(\frac{1-p_i}{N-1}\right) . \end{aligned} \quad (7)$$

The second line follows from the transformation definition in the first line by simple algebra. It portrays an intriguing result.

**Result 5.** The extropy\* of a distribution with mass function  $\mathbf{p}_N$  equals the entropy of a complementary distribution with mass function  $\mathbf{q}_N$  whose components are defined relative to those of  $\mathbf{p}_N$  by  $q_i = (1 - p_i)/(N - 1)$  for each  $i = 1, \dots, N$ . The complementary mass function  $\mathbf{q}_N$  plays the role of a photographic negative with respect to the direct mass function  $\mathbf{p}_N$ .

When  $N = 2$ , this general definition of a complementary distribution reduces to the standard definition of the distribution for an event complementary to  $E$ ,  $\tilde{E} = 1 - E$ . However, the definition generalises the concept of complementary distributions beyond events to general quantities. The content of this complementary distribution (defined only formally in terms of the transformation of the original p.m.f.) might be thought of as the distribution of “unlikeness” as opposed to the distribution of “likeliness”. The fact that an extropy\* equals some complementary entropy underlies the intimate dual relationship between extropy and entropy.

The mapping of a probability mass function  $\mathbf{p}_N$  to its complement  $\mathbf{q}_N$  is a contraction mapping. Every mass function in a unit-simplex is mapped onto a complementary function lying within an inscribed simplex of the same dimension. In turn, this complementary mass function has its own complementary distribution within a simplex inscribed in that one. The fixed-point theorem for contraction mappings assures that the uniform distribution in the centre of the unit-simplex is the unique mass function whose complementary mass function is itself. Figure 4 displays the way this contraction works in two dimensions for mass functions  $\mathbf{p}_3$ .

As a numerical and geometrical example, consider again Figure 3 in the context of the following numerical computations. Notice to begin that  $H(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) = 1.0397$  and  $J(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) = .7781$  are identifiable as points on specific entropy and extropy contours. The extropy measure would be rescaled as extropy\* via  $J^*(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) = \frac{1}{2} J(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) + \log(2) = 1.0822$  according to (7). Notice also that  $J^*(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$  also equals  $H(\frac{3}{8}, \frac{1}{4}, \frac{3}{8})$  according to Result 5. Now viewing again Figure 3Right, notice firstly that the isoextropy contour including  $J(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) = .7781$  is precisely the flipped image of the isoentropy contour including  $H(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) = 1.0397$  which appears in Figure 3Left. Both of these contours lie within and tangent to the triangle  $S_c$  if it would be inscribed within the unit-simplexes  $\mathbf{S}^2$  of Figures 3Left and 3Right. Now the numerical value of  $J^*(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$  equals the entropy value of a complementary mass function, lying on the contour including  $H(\frac{3}{8}, \frac{1}{4}, \frac{3}{8}) = 1.0822$ . This contour lies within and tangent to the subsimplex  $S_{cc}$  if it would be inscribed within  $S_c$  in Figure 3Left. This visualisation completes the understanding

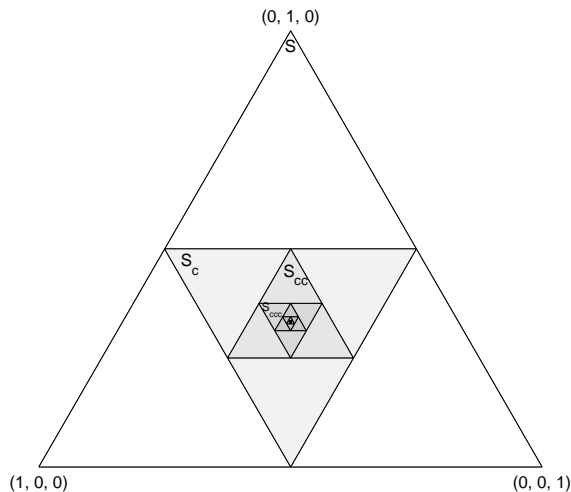


Figure 4: The complementary distribution mapping contracts the unit-simplex  $\mathbf{S}$  into the inscribed simplex  $\mathbf{S}_c$ , which it contracts in turn into the inscribed  $\mathbf{S}_{cc}$ , and then into  $\mathbf{S}_{ccc}$ , and so on.

of extropy as the dual of entropy: their iso-function contours are flips of one another; and their function values are related through the complementary contraction mapping. Notice that the symmetrical equations relating  $J(\cdot)$  to  $H(\cdot)$  in Result 4 hold for  $J^*(\cdot)$  and  $H(\cdot)$  as well.

## 6 Concluding Discussion

What’s in a name? We are aware of prior uses of the word “extropy”, documented in both the *Online Oxford English Dictionary* and in *Wikipedia*. In one usage it seems to have arisen as a metaphorical term rather than a technical term, naming a proposed primal generative natural force that stimulates order in both physical and informational systems rather than disorder. In the other, within a technical context, “extropy” also has had some parlance using it equivalently to the more commonly used “negentropy”. Neither type of usage appears to be very common. While we are not stuck on this particular word, the information measure we have introduced in this article seems aptly to merit the coinage of “extropy”. Whereas entropy is recognised as the expected log probability of the occurring value of  $X$  (a measure which could be considered “interior” to the observation  $X$ ), our proposed extropy is derived from the sum of log non-occurrence probabilities less the expected log non-occurrence probability. This could be considered to be a measure exterior to the observation  $X$ . The exterior measure of all the non-occurring quantity possibilities is complementary to the entropy measure of the unique occurring possibility. Aside from our interest in a discussion of the propriety of this name, we

are much more interested in the information-technical community's response to the content of the results we have presented.

It may be recognised that the assessment of entropies is fundamentally related to the theory of proper scoring rules for alternative forecast distributions. See, for examples, Lad (1996, ch. 6) and Gneiting and Raftery (2008). The log probability for the observed outcome of  $X$  is a proper scoring rule for forecasting mass functions with its own touted unique characteristics. The expectation of this log score is the negentropy in the distribution. Our recognition that the any extropy\* is also the entropy of a complementary distribution raises questions about the uniqueness characteristics of the log scoring rule. A detailed discussion in the context of a statistical application is in preparation.

Given the broad range of applications of entropy over the past half-century, we suspect that the awareness of extropy as a complementary dual measure to entropy will raise as many new interesting questions as it answers.

## Acknowledgements

Thanks to Patrizio Frederic, Gianfranco Lovison, Marcello Chiodi, Jim Dickey, James O'Malley and Mary Lou Zeeman for helpful comments. This research was supported by a grant from the University of Palermo dated 2007.

## References

- Gini, C.** (1912) Variabilità e mutabilità, *Studi economico-giuridici della Facoltà di Giurisprudenza dell'Università di Cagliari*, anno III parte 2<sup>a</sup>.
- Gini, C.** (1939) *Memorie di metodologia statistica, Vol. I, Variabilità e concentrazione*, Milano: Giuffrè.
- Gneiting, T., and Raftery, A.** (2007) Strictly proper scoring rules, prediction and estimation, *J. Am. Statist. Ass.*, **102**, 359-378.
- Good, I.J.** (1979) A.M. Turing's statistical work in World War II, *Biometrika*, **66**, 393-396.
- Jaynes, E.T.** (2003) *Probability theory: the logic of science*, G.L. Bretthorst (ed.), Cambridge:

Cambridge University Press.

**Kolmogorov, A.N.** (1956) On the Shannon theory of information transmission in the case of continuous signals, *IRE Trans. Information Theory*, **2**, 102-108.

**Lad, F.** (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.

**Renyi, A.** (1961) On measures of entropy and information, *Proc. Fourth Berkeley Symp. Mathematical Statistics and Probability*, Vol. I, Berkeley: University of California Press, 498-512.

**Shannon, C.E.** (1948) A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**, 379-423, 623-653; reprint, (1949) *The Mathematical Theory of Communication*, C.E. Shannon and W. Weaver (eds.), Urbana: University of Illinois Press.

## A Appendix

### Entropy $\geq$ Extropy

Let  $X$  be a random quantity with a finite discrete realm of possibilities  $\{x_1, x_2, \dots, x_N\}$  with probability masses  $p_i$ , with  $p_i = P(X = x_i)$ ,  $i = 1, \dots, N$ . We recall that

$$H(X) = - \sum_{i=1}^N p_i \log(p_i)$$

and

$$J(X) = - \sum_{i=1}^N (1 - p_i) \log(1 - p_i).$$

We introduce the following real functions defined on  $[0, 1]$  which will be useful later:

- $s(p) = -p \log(p)$  , with  $s(0) \equiv 0$  ;
- $t(p) = s(1 - p) = -(1 - p) \log(1 - p)$  , with  $t(1) \equiv 0$  ;
- $u(p) = s(p) - t(p) = -p \log(p) + (1 - p) \log(1 - p)$  , with  $u(0) \equiv u(1) \equiv 0$  .

The function  $u(p)$  satisfies the following properties (see Figure 5):

1.  $u(p) = 0$  iff  $[p = 0, \text{ or } p = 1 \text{ or } p = \frac{1}{2}]$ .
2.  $u(p) > 0$  iff  $0 < p < \frac{1}{2}$ .
3.  $u(p) < 0$  iff  $\frac{1}{2} < p < 1$ .
4.  $u(1 - p) = -u(p)$ , for all  $p \in [0, 1]$ .
5.  $u(p)$  is strictly concave in  $[0, \frac{1}{2}]$ , that is, for any given pair  $(p_1, p_2)$  with  $0 \leq p_1 < p_2 \in (0, \frac{1}{2}]$ , and for any given  $\alpha \in (0, 1)$ , we have

$$u(\alpha p_1 + (1 - \alpha) p_2) > \alpha u(p_1) + (1 - \alpha) u(p_2).$$

By exploiting the function  $u(p)$ , it is evident that

$$H(X) - J(X) = \sum_{i=1}^N u(p_i).$$

This difference is permutation invariant with respect to the components  $p_i$ .

We observe that for any  $N > 1$ , if there exist  $i \in \{1, 2, \dots, N\}$  such that  $p_i = 0$ , then by considering an arbitrary quantity  $Y$  with a realm of cardinality  $N - 1$  and probability masses  $(p_1, p_2, \dots, p_{i-1}, p_{i+1}, \dots, p_N)$  we are ensured that

$$H(X) = H(Y) \text{ and } J(X) = J(Y).$$

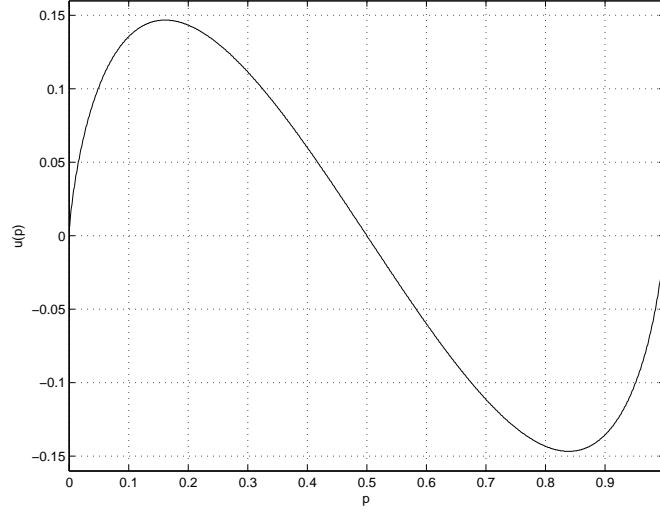


Figure 5: The function  $u(p) = -p \log(p) + (1-p) \log(1-p)$ .

We have the following result.

**Theorem 1.** Let  $X$  be a finite random quantity, with realm  $\{x_1, x_2, \dots, x_N\}$ , and probability masses  $(p_1, p_2, \dots, p_N)$  such that  $p_i > 0$ , for all  $i = 1, 2, \dots, N$ , we have

- a)  $H(X) = J(X)$  if  $N \leq 2$ ;
- b)  $H(X) > J(X)$  if  $N > 2$ .

*Proof.* Case a). If  $N = 1$  we trivially have  $H(X) = J(X) = 0$  and, if  $N = 2$  it is  $H(X) = J(X) = -x \log(x) - (1-x) \log(1-x)$ .

Case b). We distinguish two alternatives: b1)  $p_i \leq \frac{1}{2}$ ,  $i = 1, 2, \dots, N$ ; and b2)  $p_i > \frac{1}{2}$  for only one index  $i$ .

Case b1). By the hypotheses, for each  $i$ ,  $0 < p_i \leq \frac{1}{2}$  and  $\sum_{i=1}^N p_i = 1$ . It follows from Properties 1 and 2 of the function  $u(p)$  that

$$H(X) - J(X) = \sum_{i=1}^N u(p_i) > 0.$$

Case b2). To begin, suppose that  $N = 3$ . Without loss of generality we can assume  $p_3 > \frac{1}{2}$ , because of the permutation invariance of  $u(\cdot)$ ; consequently  $0 < p_1 + p_2 < \frac{1}{2}$ . Now from Property 4 we deduce

$$u(p_3) = -u(1 - p_3) = -u(p_1 + p_2).$$

Then statement  $H(X) - J(X) = u(p_1) + u(p_2) - u(p_1 + p_2) > 0$  amounts to  $u(p_1) + u(p_2) > u(p_1 + p_2)$ . Since  $u(p)$  is strictly concave over the interval  $[0, \frac{1}{2}]$  (see Property 5) and  $u(0) = 0$



we have

$$\begin{aligned}
u(p_1) &= u\left(\frac{p_2}{p_1+p_2}0 + \frac{p_1}{p_1+p_2}(p_1+p_2)\right) > \frac{p_2}{p_1+p_2}u(0) + \frac{p_1}{p_1+p_2}u(p_1+p_2) = \\
&= \frac{p_1}{p_1+p_2}u(p_1+p_2)
\end{aligned} \tag{8}$$

and

$$u(p_2) > \frac{p_1}{p_1+p_2}u(0) + \frac{p_2}{p_1+p_2}u(p_1+p_2) = \frac{p_2}{p_1+p_2}u(p_1+p_2). \tag{9}$$

From (8) and (9) it follows  $u(p_1) + u(p_2) > u(p_1+p_2)$  and then  $H(X) - J(X) > 0$ .

Generally, let  $N > 2$ . Again without loss of generality we can assume  $p_N > \frac{1}{2}$ . We have

$$u(p_N) = -u(1-p_N) = -u(p_1+p_2+\dots+p_{N-1}).$$

For each  $i = 1, \dots, N-1$ , it is easy to see that

$$u(p_i) > \frac{p_i}{p_1+p_2+\dots+p_{N-1}}u(p_1+p_2+\dots+p_{N-1}), \tag{10}$$

because of the concavity of  $u(\cdot)$ .

Finally, we have

$$H(X) - J(X) = \sum_{i=1}^N u(p_i) = \sum_{i=1}^{N-1} u(p_i) - u(p_1+p_2+\dots+p_{N-1}) > 0.$$

□